Purdue University
Purdue e-Pubs

2011 Symposium on Data-Driven Approaches to Droughts

Drought Research Initiative Network

6-22-2011

Drought Analysis Based on Copulas

Lu Chen
Texas A & M University - College Station, chl8505@126.com

Vijay P. Singh
Texas A & M University - College Station, vsingh@tamu.edu

Shenglian Guo
Wuhan University, slguo@whu.edu.cn

Drought Analysis Based on Copulas

Lu Chen[1]; Vijay P. Singh[2]; and Shenglian Guo[3]

ABSTRACT

Droughts produce a complex set of negative economic, environmental, and social impacts across a country or region. Using monthly standardized Precipitation Index (SPI) values, drought characteristics, namely, drought duration, severity, interval time and minimum SPI values, were determined. Two exponential distributions were used to model drought duration and interval time, respectively; gamma distribution was used to model for drought severity; and generalized Pareto distribution to model minimum SPI value. Several copulas in the Archimedean and meta-elliptical families were applied to construct four-dimensional joint distributions. The upstream Han River basin was selected as an example to illustrate the copulas. Results indicates that the Student copula was more appropriate for drought analysis in the selected area. Drought probabilities and return periods were calculated and analyzed based on the four-dimensional copula.

Keywords

Drought characterization; multivariate distribution; SPI; Copula

1. INTRODUCTION

Droughts caused the greatest economic losses in China during the period of 1949-1995 (Damage Report, 1995). In fact, the Chinese civilization has been deeply plagued by droughts. During the spring of last year (2010), Yunnan, Guizhou, Guangxi, Sichuan, Shanxi, Henan, Shanxi provinces of China experienced their most serious droughts in recent decades. In addition, a noticeable severe drought also occurred in China this year, when rainfall was more than 30 percent below normal since October across the five northern provinces that account for about two-thirds of Chinese wheat production. Thus, droughts are of great importance in the planning and management of water resources (Mishra & Singh, 2010).

Various indices have been developed to detect and monitor droughts, and there is extensive literature on modeling of droughts using these indices, as well as stochastic and water balance simulation models (Palmer 1965; Lana *et al*., 1998; Mishra, *et al*., 2009). Palmer Drought Severity Index (PDSI) and the standardized precipitation index (SPI) are more commonly used indices (Mishra, *et al*., 2009). Palmer (1965) proposed a moisture index (Palmer Drought Severity Index, PDSI) based on water budget accounting using precipitation and temperature data. McKee *et al.* (1993) proposed the concept of standardized precipitation index (SPI) based on the long-term precipitation or stream flow record for a desired period. PDSI has several limitations (see Alley, 1984; Guttman, 1991, 1998). For instance, the soundness of proposed water balance model is questionable, the temporal scale of PDSI is not clear, and the values of PDSI possess neither a *physical* (such as required rainfall depth) nor statistical meaning (such as recurrence probability) (Kao & Govindaraju, 2010). Due to the limitations of PDSI, Guttman (1998) recommended that the SPI can be used as the primary drought index because it is simple, spatially invariant in its interpretation, and probabilistic so that it can be used in risk and decision analysis. Therefore, SPI series was used for this study.

Drought properties are usually investigated separately by unviariate frequency anlaysis (e.g., Tallaksen *et al*., 1997, Fernández & Salas, 1999; and Cancelliere & Salas, 2004; Serinaldi, *et al*., 2010). Since droughts are complex phenomena, one variable cannot provide a comprehensive evaluation of droughts (Shiau *et al*., 2007). Separate analysis of drought duration distribution and drought severity distribution cannot reveal the significant correlation between them. Instead of using traditional univariate analysis for drought assessment, a better approach for describing drought characteristics is to derive the joint distribution of drought variables. For example, Shiau and Shen (2001), Bonaccorso *et al.* (2003), Kim *et al.* (2003), González and Valdés (2003), Salas *et al.* (2005) and Cancelliere & Salas (2010) have proposed different methods to investigate the joint distribution of drought duration and drought severity or intensity. The drawbacks of bivariate distributions are the complex mathematical derivations needed for fitting parameters from observed or generated data (Shiau, 2006).

Multivariate distributions using copulas, however, can overcome such difficulties. In recent years copulas have been used for multivariate hydrological analysis. For example, they have been used for rainfall frequency analysis (de Michele & Salvadori, 2003; Grimaldi & Serinaldi, 2006; Kao & Govindaraju, 2007; Zhang & Singh, 2007a; and Kuhn *et al*., 2007), flood frequency analysis (Favre *et al*., 2004; Shiau *et al*., 2006, Zhang & Singh, 2006; Renard & Lang, 2007; Cheng *et al*. 2009), drought frequency analysis (Shiau, 2006; Kao & Govindaraju, 2010; Song & Singh, 2010a,b), rainfall and flood frequency analysis (Singh & Zhang, 2007; Xiao *et al*. 2008; Keef *et al*., 2009; Wang, *et al*., 2010), sea storm analysis (De Michele *et al*., 2007) and some other theoretical analyses of multivariate extreme problems (Salvadori *et al*. 2007; Salvadori & de Michele, 2010). Details of the theoretical background and the use of copulas can be found in Nelsen (2006) and Salvadori *et al.* (2007).

For drought frequency analysis, Shiau (2006, 2007, 2009) modeled the joint distribution drought duration and severity using two-dimensional copulas. Mirakbari *et al.* (2010) used bivariate copula functions for regional drought analysis. The use of multivariate copulas (greater than two variables) has also emerged recently. Song and Singh (2010a) modeled the joint probability

[1] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China; Dept. of Biology and Agricultural Engineering, Texas A&M Univ., College Station, TX 77843-2117, USA. E-mail: chl8505@126.com

[2]Dept. of Biological and Agricultural Engineering & Dept. of Civil and Environmental Engineering, Texas A&M University, TAMU, College Station, TX 77843-2117, USA. E-mail: vsingh@tamu.edu

[3]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China.

E-mail: slguo@whu.edu.cn

distribution of drought duration, severity and inter-arrival time using a trivariate Plackett copula. Song and Singh (2010b) applied several meta-elliptical copulas, Gumbel-Hougaard, Ali-Mikhail-Haq, Frank and Clayton copulas to build a trivariate joint distribution for drought duration, severity and interval time, and the best-fit copula for trivariate drought analysis was selected. Serinaldi *et al.* (2009) investigated four drought characteristics, including drought length, mean and minimum SPI values, and drought mean areal extent, and built the corresponding four-dimensional joint distribution of them. Kao and Govindaraju (2010) proposed a new drought indicator using multivariate empirical copulas to compute a probability-based overall water deficit index from multiple drought-related quantities (or indices).

Until now most of the work has focused on bivariate cases. Investigators have used many different ways to build bivariate distributions of drought duration and severity. Actually, drought events have some other characteristics, such as maximum drought value corresponding to the minimum SPI (values), and drought interval time, which are mutually correlated. The studies mentioned above have only included some of the drought characteristics. None of them have taken all the characteristics of droughts mentioned above into account. However, it is important for design engineers and water resources planners to know not only the frequency of droughts but also the risk of having droughts of differing duration, severity, interval time and maximum drought degree (corresponding to the minimum SPI value) within a drought period. For this purpose, a multivariate distribution needs to be built. In order to simplify inference procedures and to derive flexible multivariate distributions, copulas can be efficiently employed.

The objective of this paper is therefore to employ the Archimedean and meta-elliptical copulas to construct four-dimensional joint distributions. The drought risk has been defined and analyzed based on the return period (recurrence interval) of drought events, which has become a standard practice for risk-based design of hydraulic structures.

2. DEFINITION OF DROUGHT AND UNIVARIATE VARIABLE

Drought identification based on an SPI series can be carried out by assuming a drought period as a consecutive number of time intervals where SPI values are less than 0 (Shiau, 2006). Important parameters for characterizing a drought used here are frequency, duration (D_d), severity (S_d), minimum SPI (MSPI) values (I_d) and interval time (L_d). Definitions of D_d and S_d can be found in Shiau (2006) and Mishra and Singh (2010). The drought interval time L_d is defined as the period elapsing from the initiation of a drought to the beginning of the next drought (Song & Singh, 2010). The minimum SPI value I_d is defined as the minimum SPI value within a drought period (Serinaldi *et al.* 2010).

Generally, drought duration is fitted as a geometric distribution (Kendall & Dracup, 1992; Mathier *et al.*, 1992) if it is treated as a discrete random variable, and an exponential distribution (Zelenhastic and Salvai, 1987) if treated as a continous random variable (Shiau, 2006). As Sklar's theorem requires the continuity of marginal distributions, the exponential distribution was selected in this study. The cumulative exponential distribution function is expressed as:

$$F_{D_d}(D_d) = 1 - \exp(-\frac{D_d - u}{\sigma}) \tag{1}$$

where u and σ are the shape and scale parameters of the exponential distribution, respectively, and $u > 0, \sigma > 0$.

The gamma distribution has generally been used to describe drought severity (Shiau, 2006), which was also selected here for fitting the drought severity distribution. The form of gamma distribution is defined as:

$$F_{S_d}(S_d) = \int_0^s \frac{S_d^{\ \alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{S_d}{\beta}} dS_d \tag{2}$$

where α and β are shape and scale parameters, respectively, and $\Gamma(\,)$ is the gamma function.

According to the data in the Han River, the generalized Pareto distribution was selected for fitting MSPI:

$$F_{I_d}(I_d) = 1 - [1 - k\left(\frac{I_d - u}{\sigma}\right)]^{1/k} \tag{3}$$

where k is the shape parameter; σ is the scale parameter; and u is the location parameter.

Shiau and Shen (2001) computed drought interval time as equal to the sum of drought duration and non drought duration, on the assumption that the drought and non drought duration follow a geometric distribution. Due to the limitations of using a discrete marginal distribution, the exponential distribution was used for fitting drought interval time. The expression of exponential distribution is given in equation (1).

3. COPULAS FOR MULTIVARIATE DISTRIBUTIONS

The problem of specifying a probability model for dependent multivariate observations can be simplified by expressing the corresponding n dimensional joint cumulative distribution (Salvadori & Michele, 2010). Following Sklar (1959) and Nelsen (2006), if $F_{1,\ 2,\ \ldots,n}(x_1, x_2, \ldots, x_n)$ is a multivariate distribution function of n correlated random variables of $X_1, X_2, \ldots, X_n$ with respective marginal distributions $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$, then it is possible to write an n-dimensional cdf with univariate margins, $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$, as follows:

$$\begin{aligned} H(x_1, x_2, \cdots, x_n) &= C(F_1(x_1), F_2(x_2), \cdots F_n(x_n)) \\ &= C(u_1, \cdots, u_n) \end{aligned} \tag{4}$$

where $F_k(x_k) = u_k$ for k=1,…, n, with $U_k \sim U(0,1)$.

Previous studies have indicated that copulas perform well for bivariate problems, and in particular, several families of Archimedean copulas, including Gumbel-Hougaard, Frank, and Clayton, have been popular choices for dependence models because of their simplicity and generation properties (Nelson, 2006). For multi-variables (greater than two), the symmetric Archimedean copula has only one parameter, which forces that all pairs of variables share the same dependence structure. In order to model different dependence structures, Grimaldi and Serinaldi (2006) applied nested classes of the Archimedean copulas. But these copulas can only model n-1 dependence. Therefore, Archimedean copulas are not adequate for modeling the dependence of three or more variables, given that different pairs exhibit widely varying degrees of dependence (Genest *et al.*, 2007). On the contrary, meta-elliptical copulas, which are extensions of the multivariate Gaussian distribution, can model arbitrary pair-wise dependencies between variables though a correlation matrix (Kao and Govindaraju, 2008). All these copulas mentioned above were used and compared in this study.

The four-dimensional symmetric and asymmetric Archimedean and meta-elliptical copulas, used in this study, are listed below:

1. Gumbel-Hougaard

The asymmetric one is given as:

$$C(u_1,u_2,u_3,u_4)=\exp\{-[(-\ln u_4)^{\theta_1}+(((-\ln u_1)^{\theta_3}+(-\ln u_2)^{\theta_3})^{\frac{\theta_2}{\theta_3}}+(-\ln u_3)^{\theta_2})^{\frac{\theta_1}{\theta_2}}]^{\frac{1}{\theta_1}}\}$$

$$1\le\theta_1\le\theta_2\le\theta_3 \tag{5}$$

The symmetric one is given as:

$$C(u_1,u_2,u_3,u_4)=\exp(-[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}+(-\ln u_3)^{\theta}+(-\ln u_4)^{\theta}]^{\frac{1}{\theta}})$$

$$\theta\ge 1 \tag{6}$$

2. Frank

The asymmetric one is given as:

$$C(u_1,u_2,u_3,u_4)=\frac{-1}{\theta_1}\ln(1+\frac{1}{e^{-\theta_1}-1}(e^{-\theta_1 u_4}-1)\cdot((1-\frac{1}{(1-e^{-\theta_2})}\cdot(1-(1-\frac{1}{(1-e^{-\theta_3})}\cdot(1-e^{-\theta_3 u_1})(1-e^{-\theta_3 u_2}))^{\frac{\theta_2}{\theta_3}})\cdot(1-e^{-\theta_2 u_3}))^{\frac{\theta_1}{\theta_2}}-1))$$

$$0\le\theta_1\le\theta_2\le\theta_3 \tag{7}$$

The symmetric one is given as:

$$C(u_1,u_2,u_3,u_4)=\frac{-1}{\theta_1}\ln(1+\frac{\prod_{i=1}^{4}(e^{-\theta_1 u_i}-1)}{(e^{-\theta_1}-1)^3}),\theta\neq 0 \tag{8}$$

3. Clayton

The asymmetric one is given as:

$$C(u_1,u_2,u_3,u_4)=(u_4^{-\theta_1}+((u_1^{-\theta_3}+u_2^{-\theta_3}-1)^{\frac{\theta_2}{\theta_3}}+u_3^{-\theta_2}-1)^{\frac{\theta_1}{\theta_2}}-1)^{\frac{-1}{\theta_1}}\quad 0\le\theta_1\le\theta_2\le\theta_3 \tag{9}$$

The symmetric one is given as:

$$C(u_1,u_2,u_3,u_4)=(u_1^{-\theta}+u_2^{-\theta}+u_3^{-\theta}+u_4^{-\theta}-3)^{\frac{-1}{\theta}}$$

$$\theta\ge 0 \tag{10}$$

4. Normal copula

$$C_N(\mathbf{u};\boldsymbol{\Sigma})=\int_{-\infty}^{\phi^{-1}(u_1)}\cdots\int_{-\infty}^{\phi^{-1}(u_n)}\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp(-\frac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x})d\mathbf{x} \tag{11}$$

where ϕ^{-1} denotes the quantile function of a standard univariate normal distribution, and $\boldsymbol{\Sigma}$ is the correlation matrix.

5. Student copula

$$C_t(\mathbf{u};\boldsymbol{\Sigma},\nu)=\int_{-\infty}^{t^{-1}(u_1)}\cdots\int_{-\infty}^{t^{-1}(u_n)}\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^n|\Sigma|}}(1+\frac{\mathbf{x}'\Sigma^{-1}\mathbf{x}}{\nu})^{\frac{\nu+n}{2}}d\mathbf{x} \tag{12}$$

where t^{-1} denotes the quantile function of a standard univariate Student t_v function; and ν denotes the degrees of freedom.

4. RETURN PERIOD FOR DROUGHT EVENTS

A common approach used in hydraulic and hydrologic design is based on frequency analysis or the recurrence interval or return period of hydrologic events (Shiau & shen, 2001). In particular, estimation of drought return periods can provide useful information for a proper water use under drought conditions (Serinaldi *et al.*, 2009). The return period of a drought can be defined as the average elapsed time or mean interval time between occurrences of critical events (Shiau and Shen, 2001; Serinaldi *et al.*, 2009). In this study, return periods of the univariate and multivariate drought events were calculated and analyzed.

4.1 Univariate return period

Shiau and Shen (2001) calculated the return period of a drought event with severity equal to or greater than a certain value *ds*. Shiau (2006) calculated the return period of a drought event with duration equal to or greater than a certain value *dl*. Similarly, the return period for drought intensity can be obtained using the same formula expressed as:

$$T_{DL}=\frac{E(L_d)}{1-F_{D_d}(d)}\;;T_{DS}=\frac{E(L_d)}{1-F_{DS}(s_d)}\;;T_{DI}=\frac{E(L_d)}{1-F_{DI}(i_d)} \tag{13}$$

where $E(L_d)$ is the expected drought interval time. Assuming the drought interval time obeying the exponential distribution in section 2, the expected value of the exponential distribution with two parameters is the sum of u and σ. Hence, $E(L_d)$ can be derived directly.

4.2 Multivariate return period

Salas *et al.* (2005) extend equation (13) to a more general case of drought events defined in terms of either severity or MSPI and duration. The interval time between two drought events E is $T_E=\sum_{j=1}^{N_d}L_{d_j}$, where L_{d_j} is the interval time between any two droughts in general (i.e., droughts not necessarily characterized by E); and N_d is the number of droughts until the next drought event E occurs. Then, the return period T is the expected value of T_E, and can be expressed as:

$$T=E(T_E)=E(N_d)E(L_d) \tag{14}$$

where $E(N_d)$ =1/$P(E)$. The multivariate return period can be calculated based on equation (14).

Shiau (2006) defined the bivariate joint return period T_{and} and T_{or} as:

$$T_{and}=\frac{E(L_d)}{1-P(X_i\ge x_i,X_j\ge x_j)}=\frac{E(L_d)}{1-F(x_i)-F(x_j)+C(F(x_i),F(x_j))} \tag{15}$$

$$T_{\text{or}}=\frac{E(L_d)}{1-P(X_i\ge x_i\text{or }X_j\ge x_j)}=\frac{E(L_d)}{1-C(F(x_i),F(x_j))} \tag{16}$$

According to equation (15), the trivariate return period can be defined as:

$$T_{and} = \frac{E(L_d)}{1-P(X_i \geq x_i, X_j \geq x_j, X_k \geq x_k)} = E(DT)/(1 - F(x_i) - F(x_j) - F(x_k) + C\left(F(x_i), F(x_j)\right) + C(F(x_i), F(x_k)) + C\left(F(x_k), F(x_j)\right) - C(F(x_i), F(x_j), F(x_k)) \quad (17)$$

$$T_{\text{or}} = \frac{E(L_d)}{1-P(X_i \geq x_i \text{ or } X_j \geq x_j \text{ or } X_k \geq x_k)} = \frac{E(L_d)}{1-C(F(x_i), F(x_j), F(x_k))} \quad (18)$$

4.3 Conditional return period

Shiau (2006) defined the bivariate conditional return period as:

$$T_{x_i|x_j} = \frac{E(L_d)}{(1 - F_{x_i}(x_i) - F_{x_j}(x_j) + F(x_i, x_j))} \quad (19)$$

where $T_{x_i|x_j}$ denotes the conditional return period for X_i given $X_j \geq x_j$.

The bivariate conditional return period of drought duration, severity and MSPI were calculated in this study.

5. DATA

The data used in this study to evaluate drought characteristics are monthly rainfall data from 1951 to 2003 in the upstream Han River, China. The Han River is a left tributary of the Yangtze River with a length of 1532 km. Monthly rainfall data from seven gauge stations, including Hanzhong, Foping, Shangzhou, Shiquan, Ankang, Xixia and Yunxian, were used in this study. The average areal rainfall of this basin was calculated based on the Thiessen polygon method. The monthly SPI series was obtained and is shown in Fig. 1.

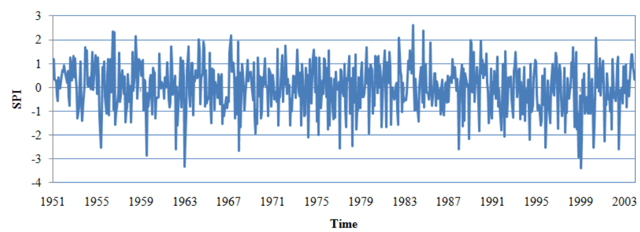


Figure 1 Monthly SPI values based on average rainfall over upstream Han River.

Genest *et al.* (2007) suggested that when looking for a copula representation of association, the most informative tool is rank scatter plots for pairs of variables against the other. Figure 2 shows the scatter plots of each variable against each other. The values of the Pearson and Kendall correlation coefficients for all drought variables are given in Table 1. Results confirmed that all variables showed positive association and a highly correlated relationship between the monthly rainfall data was observed. Serinaldi *et al.* (2009) indicated that the scatter plot matrix method used above merged information on marginal and dependence structure. They filtered out the marginal information by using the pseudo observations $u_{ij} = F_i(x_{ij})$, where i=1, ··; d and j=1, ··; m. The same method was also used in this study and is shown in Fig. 3. It is indicated from Fig. 3 that some samples seem to show accumulation of points in the lower left corner, which shows the possible lower tail dependence. The dependence analysis demonstrates that the association properties and tail dependence among drought variables must be taken into account and then the appropriate copula class needs to be selected.

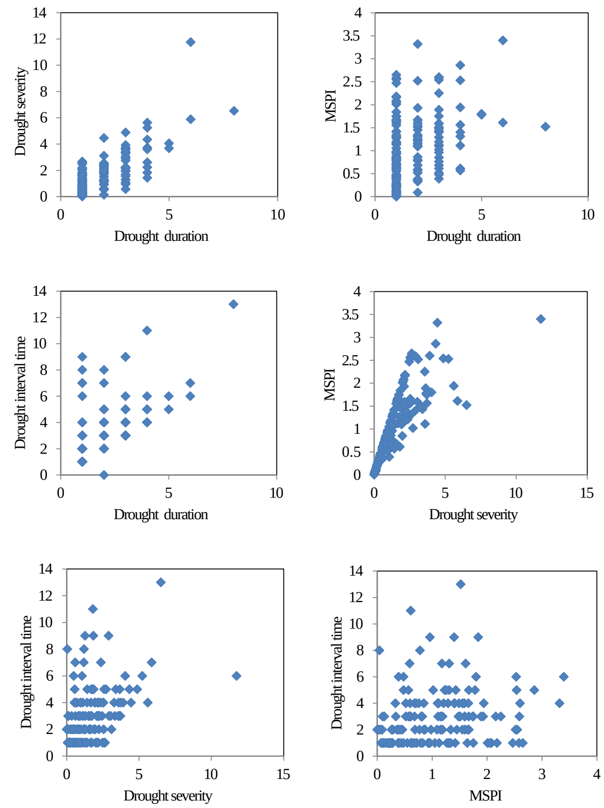


Figure 2 Scatter-plots of the pair-wise drought variables.

Table 1 Values of Pearson and Kendall correlation coefficient for all drought variables

Correlation coefficient	Duration	Severity	MSPI	Interval Time
Duration	1.000	0.749	0.360	0.650
Severity	0.526	1.000	0.792	0.451
MSPI	0.278	0.781	1.000	0.197
Interval time	0.607	0.358	0.199	1.000

6. APPLICATION

6.1 Estimation of marginal distributions

Parameters of marginal distributions were estimated by L-moments (Hosking, 1990). The parameters of the marginal distribution for drought duration were u=0.685 and σ=1.168, for drought severity were α =1.231and β =1.312, for MSPI were u=0.05, σ =1.591 and k=0.526, and or drought intensity were u=0.796, σ=2.178. Figure 4 compares computed and empirical marginal distributions of the observed drought duration, interval, severity and MSPI. It is demonstrates that the theoretical and empirical values fit well for all the marginal distributions.

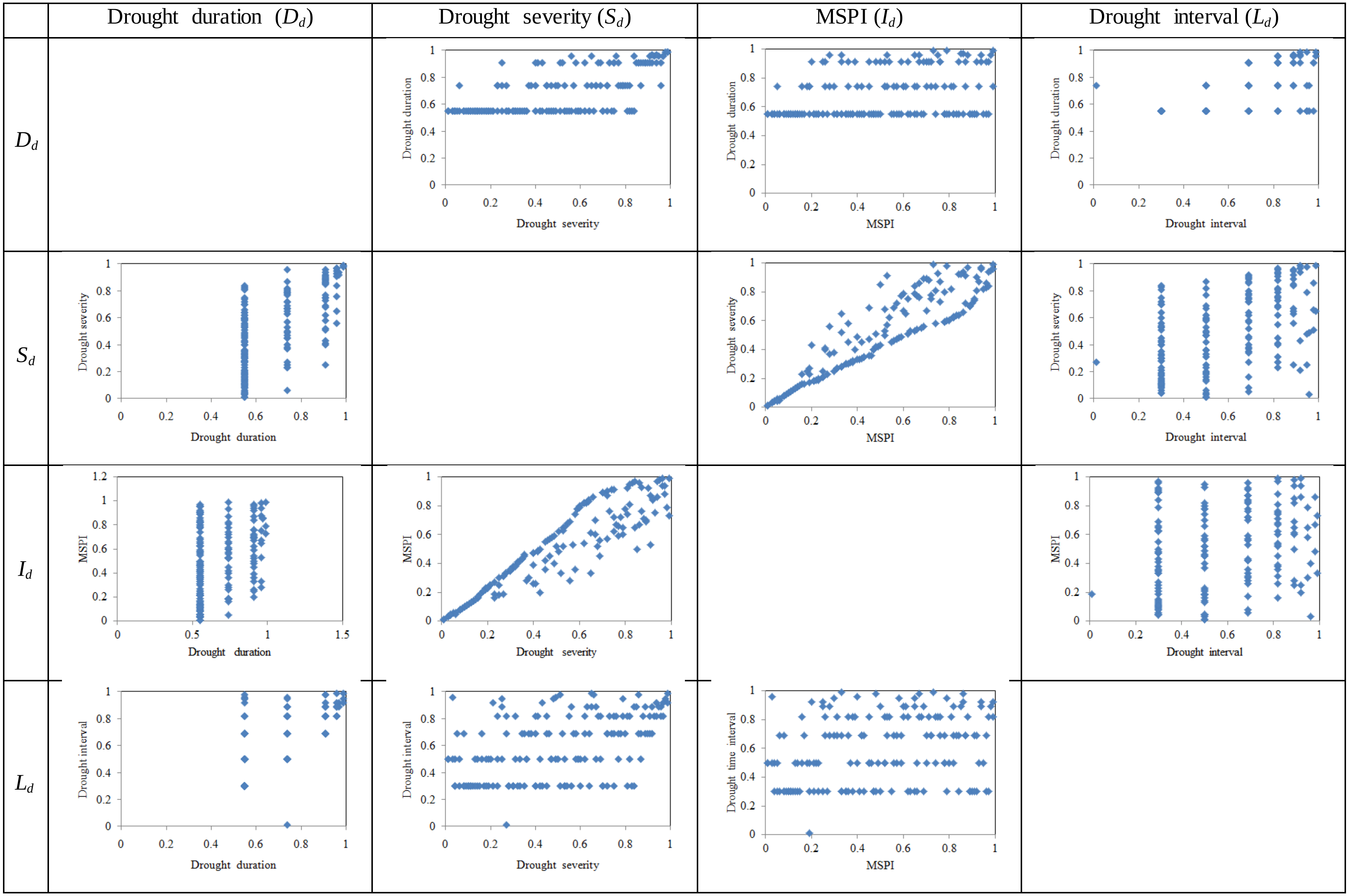
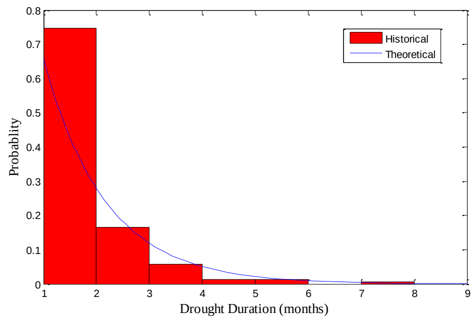
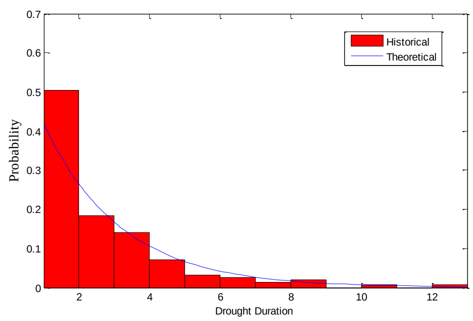


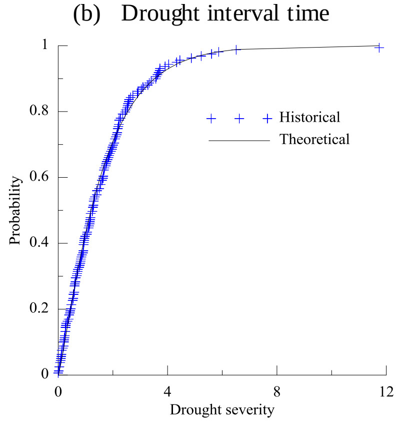
Figure 3 Scoter plots for pseudo observations



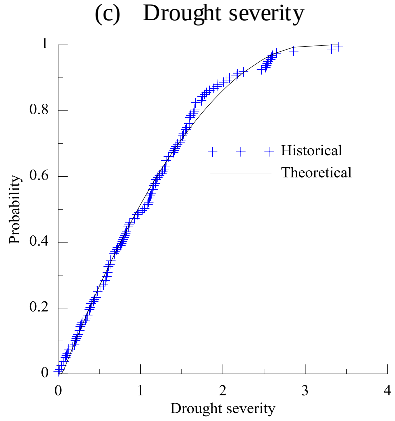
(a) Drought duration



(b) Drought interval time



(c) Drought severity



(d) MSPI

Figure 4 Frequency curves of marginal distribution

6.2 Estimation of joint distributions

Four-dimensional Archimedean and meta-elliptical copulas were tested to determine the best-fit copula for modeling the dependence amongst the four drought characteristics. For the Archimedean family, three widely used copulas, including Gumbel-Hougaard, Frank and Clayton, were used; for the meta-elliptical copulas, normal and Student copula were in use. A pseudo-likelihood technique involving the ranks of the data was used for estimating parameters of the four-variate symmetric and asymmetric Archimedean copulas. The estimated parameters of symmetric and asymmetric Archimedean copulas were given in Table 2. Both the log-pseudo likelihood and the inversion of Kendall's tau method (Genest *et al.*, 2007) were used to estimate the parameters of normal copula. The estimated parameters of normal copula were listed in Table 3. Parameters of Student copula were estimated by maximum pseudo-likelihood method, the values of which are 0.75, 0.391, 0.610, 0.811, 0.416, 0.191. The degree of freedom is 2 for the Student copula.

In order to select the appropriate copula, the root mean square error (RMSE) and the Akaike information criterion (AIC) were used (Zhang & Singh, 2006). The RMSE and AIC values of the Archimedean and meta-elliptical copulas were shown in Table 4. It is indicated that asymmetric copulas give a better fit than symmetric copulas in the Archimedean family. Generally meta-elliptical copulas fit better than the Archimedean ones, except the Frank copula. The RMSE and AIC values of Student copula is more or less the same as the Frank one. However, Fig. 2 shows that tail dependence exists in the drought variable. As Student copula can describe the tail dependence between variables, this copula was selected for the drought analysis hereafter. Fig. 5 shows the pp-plot comparing observed and theoretical values of Student copula with 2 degrees of freedom (Genest *et al.*, 2007; Serinaldi *et al.*, 2009). The observed and theoretical values fit each other well. The Genest-Rémillard goodness test based on the Cramér-von Mises S_n statistic was applied (Genest *et al.*, 2008) for the selected Student copula. The S_n value is 0.023 with the p value 0.902 obtained by the parametric bootstrapping method. The goodness-of-fit test indicates that the constructed Student copula provides a good fit.

Table 2 Estimated parameters of symmetric and asymmetric Archimedean copulas

Family	Symmetric	Asymmetric		
	θ	θ_1	θ_2	θ_3
Gumbel	1.55	1.36	1.36	2.91
Frank	4.13	2.96	2.96	12.52
Clayton	1.05	0.54	0.54	7.44

Table 3 Estimated parameters (correlation matrix) of normal copulas (the superdiagonal elements are the values based on inversion of Kendall's tau method; the subdiagonal elements are the values based on the log-pseudo likelihood method).

1.000	0.754	0.392	0.594
0.728	1.000	0.819	0.413
0.351	0.791	1.000	0.185
0.552	0.364	0.151	1.000

Table 4 RMSE and AIC values of different copulas

Family	Archimedean						Meta-elliptical	
	Gumbel		Frank		Clayton		Norm	Student
	A	B	A	B	A	B		
RMSE	0.065	0.060	0.046	0.041	0.061	0.061	0.048	0.042
AIC	-429	-442	-483	-502	-439	-439	-474	-495

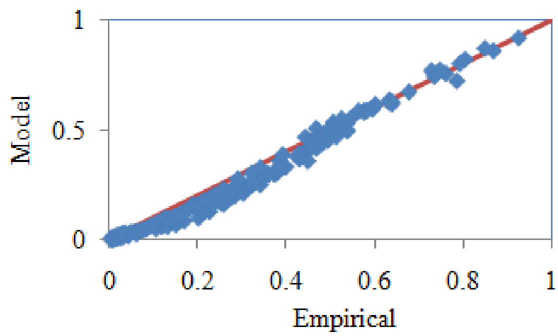


Figure 5 PP- plot of joint distribution for the four drought characteristics described in the text.

6.3 Drought probability analysis

In this study, drought events were defined by drought duration, severity, interval time and MSPI. It is necessary to know the occurrence probabilities of arbitrary drought events. Table 5 gives the joint probabilities of some drought events $E=\{D_d \le d_d,\ S_d \le s_d,\ I_d \le i_d,\ L_d \le l_d\}$.

The probability of events $E=\{D_d \le d_d,\ S_d \le s_d,\ I_d \le i_d\}$ under the condition $L_d \le l_d$ can be defined as:

$$P(D_d \le d_d, S_d \le s_d, I_d \le i_d | L_d \le l_d) = P(D_d \le d_d, S_d \le s_d, I_d \le i_d, L_d \le l_d)/P(L_d \le l_d) \quad (19)$$

Table 5 also gives the conditional probabilities of some drought events under the condition $L_d \le l_d$.

Table 5 Joint probabilities of drought characteristics

D_d	S_d	I_d	L_d	Joint probabilities	Conditional probabilities
1	1	0.3	1	0.028	0.314
2	1.5	0.6	2	0.151	0.356
3	3	1.2	3	0.401	0.630
5	6	2.4	6	0.867	0.955
8	12	3.6	12	0.994	0.9998

6.4 Return period analysis

The average drought interval time estimated from the observed data and theoretical distribution is both 3. Therefore, the calculated result 3 was used hereafter. According to equation (13), return periods of 2, 5, 10, 20 and 50 years defined by separate drought duration, severity and MSPI were given in Table 6.

According to the drought data D_d, S_d, I_d given in Table 5, the bivariate return periods T_{and} and T_{or} of either duration, severity or MSPI were calculated and given in Table 7. The trivariate return periods T_{and} and T_{or} of duration, severity or MSPI were also listed in Table 7.

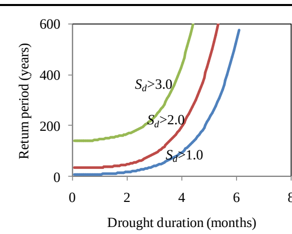
Figure 6 defines the bivariate conditional return period of drought duration, severity and MSPI by equation (19).

Table 6 Return periods defined by each drought variable

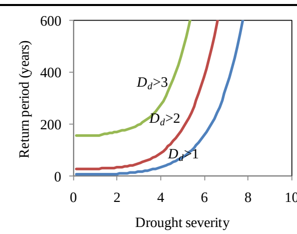
Return period (a)	D_d (months)	S_d	I_d
2	0.77	0.18	0.17
5	1.84	1.65	1.28
10	2.65	2.66	1.83
20	3.46	3.64	2.21
50	4.53	4.91	2.54

Table 7 Joint return periods (years) of the drought events *E*

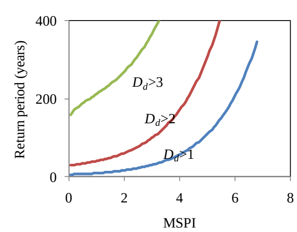
$D_d>d_d, S_d>s_d$		$D_d>d_d, I_d>i_d$		$S_d>s_d, I_d>i_d$		$D_d>d_d, S_d>s_d, I_d>i_d$	
T_{and}	T_{or}	T_{and}	T_{or}	T_{and}	T_{or}	T_{and}	T_{or}
5.6	3.7	4.3	3.2	5.4	3.5	5.7	3.2
11.7	6.2	11.6	4.0	7.8	4.2	12.9	4.0
34.3	15.1	33.7	6.6	22.8	7.1	40.2	6.6
265.0	98.0	255.0	42.1	209.8	49.5	337.9	42.0



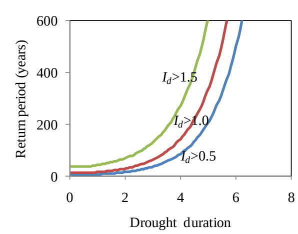
(a) $T_{D_d|S_d}$



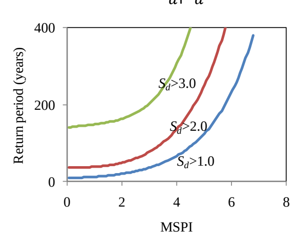
(b) $T_{S_d|D_d}$



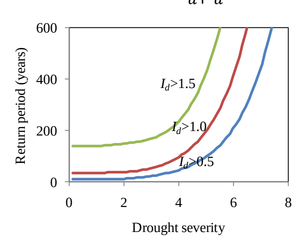
(c) $T_{I_d|D_d}$



(d) $T_{D_d|I_d}$



(e) $T_{I_d|S_d}$



(f) $T_{S_d|I_d}$

Figure 6 Conditional return periods of drought events.

7. CONCLUSIONS

In this study, a drought is defined by drought duration, severity, MSPI and interval time. The upstream Han River is selected as a case study. The exponential, gamma and generalized Pareto distributions are used to fit univariate data. The Archimedean and meta-elliptical copulas are used to establish the joint multivariate distributions. The joint probabilities and return period are then estimated. The main conclusions of this study are:

(1) The established marginal distribution of the four drought variables can fit the empirical data well, and can be used for drought analysis

(2) The RMSE and AIC values are used to select the appropriate copula. Considering both the RMSE and AIC values and the dependence structure, the Student copula is found to be the best.

(3) The drought risk is estimated based on joint probabilities and return periods, which give important information for water management and planning.

8. ACKNOWLEDGMENTS

The study is financially supported by Ministry of Science and Technology of China (2009BAC56B02) and Chinese Natural Science Fundation (51079100).

9. REFERENCES

[1] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (Nov. 1993), 795-825. DOI= http://doi.acm.org/10.1145/161468.16147.

[2] Ding, W. and Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.

[3] Fröhlich, B. and Plate, J. 2000. The cubic mouse: a new device for three-dimensional input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands, April 01 - 06, 2000).

[4] Alley, W. M. 1984. The Palmer Drought Severity Index: Limitations and assumptions. *J. Climate Appl. Meteor.,* **23,** 1100–1109.

[5] Bonaccorso, B., Cancelliere, A., and Rossi, G. 2003. An analytical formulation of return period of drought severity. *Stochastic Environmental Research and Risk Assessment* **17**(3), 157-174.

[6] Cancelliere, A. and Salas, J. D. 2010. Drought probabilities and return period for annual streamflows series. *Journal of Hydrology* 391(1-2): 77-89.

[7] Cancelliere, A. and Salas, J.D. 2004. Drought length properties for periodic-stochastic hydrological data, *Water Resources Research.* **40,** W02503 doi:10.1029/2002WR001750.

[8] Cheng Wang, Ni-Bin Chang & Gour-TsyhYeh 2009. Copula-based flood frequency (COFF) analysis at the confluences of river systems. *Hydrological Processes,* E23 (10), 1471-1486.

[9] Damage Report. 1995. *Report of the Damage Caused by Disaster in China*. China Statistical Press: Beijing.

[10] De Michele, C. & Salvadori, G. 2003. A generalized Pareto intensity duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.* 108(D2):4067. doi:10.1029/2002JD002534.

[11] De Michele, Salvadori C. G., Passoni G., & Vezzoli R. 2007. A multivariate model of sea storms using copulas, *Coastal Engineering,* 54 (10), 734-751.

[12] Favre A-C, Adlouni S., Perreault L., Thiémonge N. &Bobée B. 2004. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.*, **40**:W01101, 12.

[13] Fernández, B. & Salas, J.D. 1999. Return period and risk of hydrologic events I: mathematical formulation. *Journal of Hydrologic Engineering,* **4** (4): 297–307.

[14] Genest, C., Favre, A. C., Béliveau, J. &Jacques, C. 2007. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour. Res.* 43(9): W09401.

[15] González, J. & Valdés, J. B. 2003. Bivariate drought recurrence analysis using tree ring reconstructions, *Journal of Hydrologic Engineering ASCE* **8**(5), 247-258.

[16] Grimaldi, S & Serinaldi, F. 2006. Design hyetographs analysis with 3-copula function. *Hydrological Sciences Journal*, 51(2), 223-238.

[17] Guttman, N. B. 1991. A sensitivity analysis of the Palmer hydrologic drought index. *Journal of the American Water Resources Association* 27(5): 797-807.

[18] Guttman, N.B. 1998. Comparing the Palmer drought index and the standardized precipitation index. *J. Amer. Water Resour. Assoc.*, 34, 113-121.

[19] Hosking, J. R. M. 1990. *L*-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, **52**, 105-124.

[20] Kao S. C. & Govindaraju R. S. 2007. A bivariate frequency analysis of extreme rainfall with implications for design. *Journal of Geophysical Research,* 112, D13119, doi:10.1029/2007JD008522.

[21] Kao, S.-C. & Govindaraju, R. S. 2008. Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resour. Res.* 44(2): W02415.

[22] Kao, S.-C. & Govindaraju, R. S. 2010. A copula-based joint deficit index for droughts. *Journal of Hydrology*, 380(1-2): 121-134.

[23] Keef C., Svensson C. & Tawn J. A. 2009. Spatial dependence in extreme river flows and precipitation for Great Britain. *Journal of Hydrology*, vol. 378, 240-252.

[24] Kendall, D. R. and Dracup, J. A. 1992. On the generation of drought events using an alternating renewal-reward model, *Stochastic Hydrology and Hydraulics* **6**(1), 55–68.

[25] Kim, T. W., Valdés, J. B., and Yoo, C. 2003. Nonparametric approach for estimating return periods of droughts in arid regions, *Journal of Hydrologic Engineering ASCE* **8**(5), 237-246.

[26] Kuhn G., Khan S., Ganguly A. R. & Branstetter M. L. 2007. Geospatial–temporal dependence among weekly precipitation extremes with applications to observations and

climate model simulations in South America. *Advances in Water Resources*, 30, 2401–2423.

[27] Lana X, Serra C, and Burgueňo, A. 1998. Spatial and temporal characterization of annual extreme droughts in catalonia (Northern Spain). *Int. J. Clim*. 18: 93–110.

[28] Mathier, L., Perreault, L., Bobe, B., and Ashkar, F. 1992. The use of geometric and gamma-related distributions for frequency analysis of water deficit, *Stochastic Hydrology and Hydraulics* **6**(4), 239–254.

[29] Mirakbari, M., Ganji, A., & Fallah, S.R. 2010. Regional Bivariate Frequency Analysis of Meteorological Droughts. *Journal of Hydrologic Engineering*, 15(12), 985-1000.

[30] Mishra, A. K. & Singh, V. P. 2010. A review of drought concepts. *Journal of Hydrology* 391(1-2): 202-216.

[31] Mishra, A., Singh, V. & Desai, V. 2009. Drought characterization: a probabilistic approach. *Stochastic Environmental Research and Risk Assessment* 23(1): 41-55.

[32] Nelsen, R.B. 2006. *An Introduction to Copulas*, second edition. Springer-Verlag, New York.

[33] Palmer, W.C. 1965. *Meteorological drought*. Research paper no. 45, US Department of Commerce, Weather Bureau, Washington, DC

[34] Renard, B., & Lang, M. 2007. Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology, *Advances in Water Resources*, 30 (4), 897–912.

[35] Salas, J. D., Fu, C., Cancelliere, A., Dustin, D., Bode, D., Pineda, A., and Vincent, E., (2005) Characterizing the severity and risk of drought in the Poudre River, Colorado, *Journal of Water Resources Planning and Management ASCE* **131**(5), 383-393.

[36] Salvadori, G., De Michele, C., Kottegoda, N.T. & Rosso, R. 2007. *Extremes in nature: An approach using copulas*. Springer, 2007.

[37] Salvadori, G. & De Michele, C. 2010. Multivariate multiparameter Extreme Value models and Return Periods: a copula approach, *Water Resour. Res*., 46, W10501, doi:10.1029/2009WR009040.\

[38] Serinaldi, F., Bonaccorso, B., Cancelliere, A. & Grimaldi, S. 2009. Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth, Parts A/B/C* 34(10-12): 596-605.

[39] Shiau, J.T., Wang, H.Y. and Chang, T.T. 2006. Bivariate frequency analysis of floods using copulas. *J Am. Water Resour. Assoc*. 42(6):1549-1564.

[40] Shiau, J. 2006. Fitting Drought Duration and Severity with Two-Dimensional Copulas. *Water Resources Management* 20(5): 795-815.

[41] Shiau, J. T. & Shen, H. W. 2001. Recurrence analysis of hydrologic droughts of differing severity. *Journal of Water Resources Planning and Management-Asce* 127(1): 30-40.

[42] Shiau, J.T. & Modarres, R. 2009. Copula-based drought severity-duration-frequency analysis in Iran. *Meteorological Applications*, 16: 481–489.

[43] Shiau, J.-T., Feng, S. & Nadarajah, S. (2007) Assessment of hydrological droughts for the Yellow River, China, using copulas. *Hydrological Processes* 21(16): 2157-2163.

[44] Singh V.P., Zhang L. (2007) IDF curves using the Frank Archimedean copula. *Journal of Hydrologic Engineering*, 12:6(651):1084-0699.

[45] Sklar A. (1959) Fonctions de répartition à n dimensions et leursmarges. *Publ. Inst. Stat. Univ. Paris*, 8, 229–231.

[46] Song, S. & Singh, V. P. 2010a. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data, *Environmental Research and Risk Assessment*, 24 (3), 425-444.

[47] Song, S. & Singh, V. P. 2010b. Frequency analysis of droughts using the Plackett copula and parameter estimation by genetic algorithm, Stochastic *Environmental Research and Risk Assessment*, 24 (5), 783–805.

[48] Tallaksen, L.M., Madsen, H. & Clausen, B. (1997) On the definition and modeling of stream drought duration and deficit volume, *Hydrological Sciences Journal*. **42** (1), 15–33.

[49] Wang X., Gebremichael, M., & Yan, J. 2010. Weighted likelihood copula modeling of extreme rainfall events in Connecticut. *Journal of Hydrology*, 390(1-2), 108-115.

[50] Xiao, Y., Guo, S.L., Liu P., Yan, B.W., & Chen, L. 2009. Design Flood Hydrograph Based on Multicharacteristic

[51] Synthesis Index Method. J*ournal of Hydrologic Engineerin*g **14**(12): 1359-1364.

[52] Zelenhastic, E. and Salvai, A. 1987. A method of streamflow drought analysis, *Water Resources Research* **23**(1), 156–168.

[53] Zhang, L. & Singh, V.P. 2006. Bivariate flood frequency analysis using the copula method. *J Hydrol. Eng.* 11(2):150-164.

[54] Zhang, L. & Singh, V.P. 2007a. Gumbel–Hougaard copula for trivariate rainfall frequency analysis. *J Hydrol. Eng.* 12(4):409-419.